Deep Learning Models for Digital Medical Imaging: A Survey

Bing Liu[a] (bingliu23@mails.jlu.edu.cn), Xueju Wang[a] (xueju@jlu.edu.cn)
Yujia Cong[b] (Yujia23@mails.jlu.edu.cn), Lele Cong[c]
(congll18@mails.jlu.edu.cn), Xianling Cong[d] (congxl@jlu.edu.cn), Shisong
Tang[e] (tangss21@tsinghua.org.cn) Hechang Chen[f,g] (chenhechang01@163.com)

[a] Pathology Department,China-Japan Union Hospital of Jilin University,Changchun,130033,China

[b]Department of Dermatology,China-Japan Union Hospital of Jilin University,Changchun,130033,China

[c] Department of Neurology, China-Japan Union Hospital of Jilin University, Changchun, 130033, China

[d]Biobank, China-Japan Union Hospital of Jilin University, Changchun, 130033, China

[e] Tsinghua university, Beijing, 100084, China

[f] School of Artificial Intelligence,Jilin University,Changchun,130033,China

[g] Engineering Research Center of Knowledge-Driven Human-Machine Intelligence,Ministry of Education,Changchun,130033,China

Corresponding Author:
Xueju Wang
Pathology Department,China-Japan Union Hospital of Jilin University,Changchun,130033,China
Tel: 13504402297
Email: xueju@jlu.edu.cn

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

Lele Cong

Department of Neurology,China-Japan Union Hospital of Jilin University,Changchun,130033,China

Tel: 13634303853

Email: congll18@mails.jlu.edu.cn

Xianling Cong

Biobank,China-Japan Union Hospital of Jilin University,Changchun,130033,China

Tel: 15904420533

Email: congxl@jlu.edu.cn

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

Deep Learning Models for Digital Medical Imaging: A Survey

Bing Liu[a], Xueju Wang[a,*], Yujia Cong[b], Lele Cong[c,*], Xianling Cong[d,*], Shisong Tang[e], Hechang Chen[f,g]

[a]*Pathology Department,China-Japan Union Hospital of Jilin University,Changchun,130033,China*
[b]*Department of Dermatology,China-Japan Union Hospital of Jilin University,Changchun,130033,China*
[c]*Department of Neurology,China-Japan Union Hospital of Jilin University,Changchun,130033,China*
[d]*Biobank,China-Japan Union Hospital of Jilin University,Changchun,130033,China*
[e] *Tsinghua university, Beijing, 100084, China*
[f]*School of Artificial Intelligence,Jilin University,Changchun,130033,China*
[g]*Engineering Research Center of Knowledge-Driven Human-Machine Intelligence,Ministry of Education,Changchun,130033,China*

Abstract

Deep learning and its applications in medical imaging, especially in assisting with the diagnosis and treatment of diseases, receive considerable attention in the literature these days. When confronted with common ailments, deep learning models can provide diagnostic results at a speed far exceeding that of clinicians, while maintaining consistent accuracy. Despite the proliferation of relevant literature, existing reviews tend to be one-sided, focusing either exclusively on the medical perspective or solely on deep learning models. Limited attention has been given to the systematic integration of these two domains. This paper, endeavors to fill in this gap by offering a systematic elaboration of the convergent development between deep learning and medicine, encompassing the entire chain from data collection, through model application, to clinical tasks. By contributing this holistic perspective, the review aims to propel the boundary-

*Corresponding author.
Email addresses: bingliu23@mails.jlu.edu.cn (Bing Liu), xueju@jlu.edu.cn (Xueju Wang), Yujia23@mails.jlu.edu.cn (Yujia Cong), congll18@mails.jlu.edu.cn (Lele Cong), congxl@jlu.edu.cn (Xianling Cong), tangss21@tsinghua.org.cn (Shisong Tang), chenhechang01@163.com (Hechang Chen)

Preprint submitted to Engineering Applications Of Artificial Intelligence November 25, 2024

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

pushing advancement of deep learning models within the medical field. First, for input date, we delve into the challenges posed by data volume and feature selection, the intricacies of model preprocessing, and data fusion. Second, for deep learning models , our discussion centers on pivotal concerns such as generalization capacity, robustness, and interpretability. Then, for clinical tasks, we offer a concise overview of the three core modules: disease segmentation, diagnosis, and classification. Finally, we provide a critical discussion of the field's current state as a whole, including the identification of key challenges, open problems, and outlining promising future directions. We hope this survey will ignite further interest in the community and provide researchers with an up-to-date reference regarding applications of deep learning models in medical imaging.

Keywords: Deep learning; Medical image; Date input; Deep learning models; Clinical tasks

1. Introduction

Since the emergence of deep learning, have demonstrated remarkable improvements on many fields, including image recognition **?** , speech recognition **?**, and intelligent navigation **?**. As deep learning models undergo continuous refinement and enhancement, their application domains are expanding, and their performance is improving, thereby attracting widespread recognition of their practical significance (Figure 2)(Figure 3). In the diagnosis of diseases, relying solely on experience-based judgments by clinical experts will inevitably lead to subjective errors, particularly when confronted with a large clinical workload. However, the combination of deep learning models with medical tasks not only elevates work efficiency substantially but also aids in addressing challenges stemming from experience-based issues, thereby facilitating the smooth execution of clinical operations. Currently, deep learning models are extensively utilized in medical image segmentation, disease diagnosis, and classification tasks, particularly in pathology and medical image, where their image diagnostic capabili-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

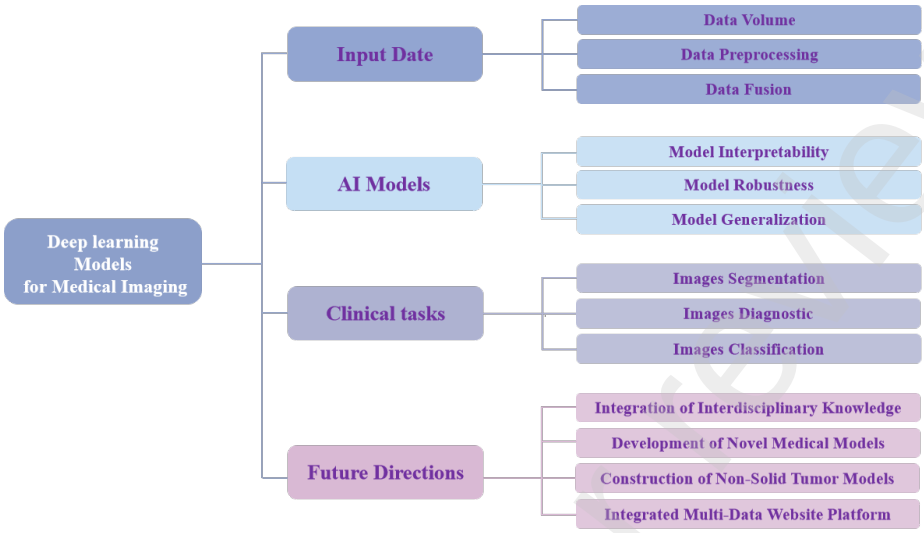

Figure 1: Framework Classification Table

ties have been comprehensively demonstrated. Additionally, deep learning has demonstrated significant potential in cutting-edge areas of disease treatment, such as neoadjuvant therapy and targeted drug therapy. Its superior computational capabilities have provided clinical research with more precise and accurate gene screening results. In summary, the integrated development of deep learning and the medical field will jointly promote the bidirectional progress of disease diagnosis and treatment, inject new vitality into the healthcare industry, and elevate medicine to new heights.

1.1. Existing Work

The evolution of medical task processing can be delineated into three distinct epochs: the knowledge-driven era, the data- driven era, and the knowledge and date co-driven era. Next, these three parts will be introduced in detail.

In the knowledge-driven era, clinical experts primarily relied on their profound medical knowledge, intricate understanding of disease mechanisms, and vast work experience to diagnose diseases. While they provided invaluable di-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

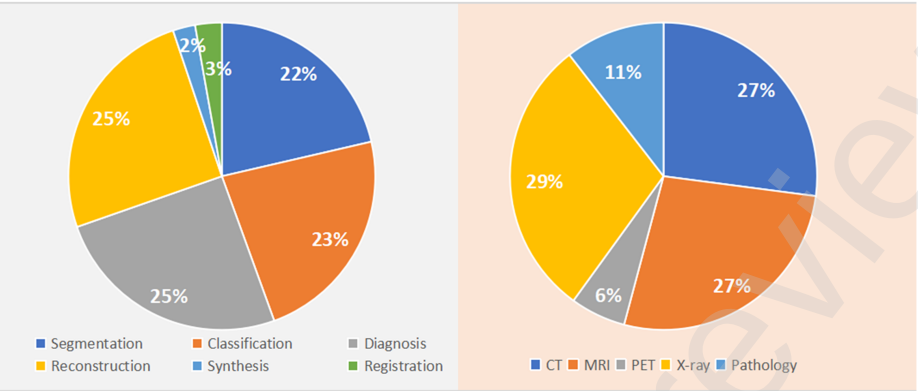

Figure 2: (Left) The pie chart shows the number of articles that applied deep learning models to different medical tasks in 2024. (Right) The pie chart displays the number of articles that utilized deep learning models in various medical imaging domains in 2024.

agnostic insights in cases of rare diseases, subjective biases were inevitable, particularly when confronted with heavy and complex workloads.

The advent of the data-driven era was marked by the rapid development of machine learning and its extensive integration into the medical field **??** . During this era, machine learning models achieved remarkable results in tasks such as disease segmentation, diagnosis, and classification, effectively solves the subjective problem of manual interpretation by experts. Especially, the introduction of deep convolu- tional neural networks by LeCun et al. in 1989 propelled the integration of artificial intelligence with the medical field to unprecedented heights. As the architecture of deep learning models continues to be optimized and enhanced, their appli- cation scope has broadened, encompassing diverse medicaldomains such as neurology, retina, lungs, digital pathology, breast, heart, and musculoskeletal systems **?**. In the data-driven era, despite the considerable convenience it has brought, numerous limitations have also surfaced **?**. Especially, machines are vulnerable to interference from noise, artifacts, and other confounding factors during data interpre- tation, and their

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

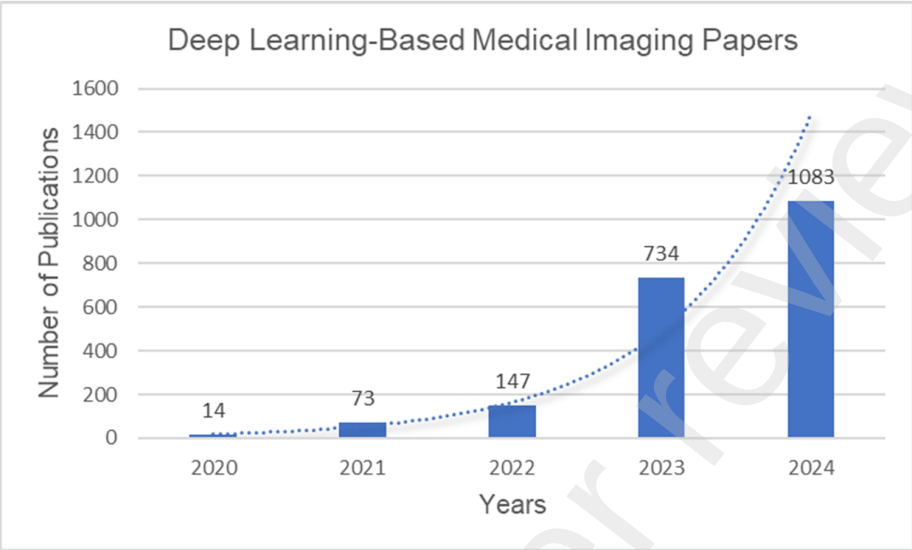

Figure 3: The chart demonstrates a consistent growth in the number of publications combining deep learning models with medical fields in recent years, from 2020 to 2024.

model performance can be elusive when con- fronted with intricate pathogenic mechanisms.

Consequently, the knowledge and date co-driven models is well-timed. By incorporating relevant domain knowledge, these models not only alleviate the workload of clinical experts but also enable the acquisition of higher-quality image input data, ultimately enhancing model performance and diagnostic efficiency. In comparison to traditional deep learning models, knowledge and data dual-drive models are more aligned with practical clinical demands.However, in the face of increasingly complex medical problems, further research is still needed to refine model mechanisms, enrich clinical task types, and continuously innovate and improve the deep learning model architecture. Through continuous optimization of processes and active integration of expert knowledge, we can effectively address diagnostic challenges in the medical field.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

1.2. Limitations

In recent years, the integrated application of deep learning models in medicine has yielded significant achievements in clinical settings. For instance, Dinggang Shen and his team detailed the specific procedures of deep learning models in medical applications in their 2017 article **?**. Similarly, in their 2020 article, Heang-Ping Chan et al. extensively outlined the application of deep learning in medical image analysis **?**. In 2022, Geert Litjens et al. published a comprehensive review on the utilization of deep learning in medicine, encompassing its application in tasks like segmen- tation, diagnosis, and classification, while outlining specific practices across diverse medical fields **?**. However, the promotion of these articles in the medical field is commonly hindered by limitations and challenges, primarily manifested in the lack of a systematic literature review and analysis that integrates deep learning models with medicine (Figure 1):

(1) From a medical perspective, existing articles have not effectively integrated deep learning models and summarized their applications from the perspectives of expert knowledge and medical tasks. The lack of a clinical perspective is bound to lead to feasibility issues in the practical application of the models, making it difficult to accurately handle complex disease relationship tasks. (2) From a model perspective, most current articles lack a comprehensive deep learning framework, namely, data input - AI models - clinical tasks . This deficiency hinders the interpretation and usage by clinicians, thereby affecting the promotion of deep learning models in practical work.

Therefore, the primary objective of this paper is to fill this research gap by comprehensively summarizing the specific applications and feasibility of deep learning models in medical applications from the three key perspectives of "input-model-output" . Through a thorough analysis of various aspects of the model architecture, this paper strives to provide a solid theoretical foundation and practical guidance for the further development and implementation of deep learning in the medical field.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

1.3. Innovations

To address the limitations of current models and promote the development of deep learning in the medical field, this paper systematically summarizes existing literature from the perspective of the deep learning framework, focusing on the "input-model-output" chain (Table 1). At the input level, this paper provides a multi-dimensional summary and induction of data volume, preprocessing procedures, and data fusion techniques. At the model level, it examines the generality, robustness, and interpretability of various models, aiming to optimize model performance. At the output level, this paper categorizes medical tasks into three basic modules: segmentation, diagnosis, and classification, and assesses current challenges and prospects. Furthermore, by analyzing deep learning models, this paper not only enhances their practical application significance but also supplements existing theoretical findings, thereby strengthening the feasibility of deep learning in medicine and advancing its application in this field.

Category	**Subcategory**	**Papers**
Input Date	Data Volume	[7][24][27][37][48][64][71][89][90][95][97]
	Data Preprocessing	[12][17][19][22][23][26][32][38][58][82][94][109]
	Data Fusion	[33][36][50][63][78]
AI Models	Model Interpretability	[18][44][45][61][79][92][103]
	Model Robustness	[29][31][35][40][60][74][84][87][98][100][110]
	Model Generalization	[13][20][30][41][46][83][91][99][101][105]
Clinical tasks	Images Segmentation	[1][25][42][56][62][72][73][85][106]
	Images Diagnostic	[10][15][34][47][57][65][104][108]
	Images Classification	[2][8][16][28][49][55][69][86][107]

Table 1: Bibliographic Classification Table

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

2. Input

In the realm of deep learning frameworks, this paper predominantly examines the input module and delves into the intricacies and challenges encountered in its application within the medical. This paper makes a comprehensive analysis of the existing models, and expounds the shortcomings of the existing models in medical application from three aspects: data quantity, preprocessing method and data fusion.

2.1. Data Volume

The progress of deep learning is heavily contingent upon the availability of large-scale data. Data serves as the cornerstone for the learning of deep learning algorithms, as well as the foundation for model development and implementation. Currently, the data utilized in medical models predominantly stems from publicly accessible databases. While this facilitates researchers' access, it concomitantly presents numerous challenges, including limited sample sizes, suboptimal data feature representation, and incomplete feature coverage. These issues profoundly influence the training efficacy of the models, thereby affecting their clinical performance. Therefore, after a thorough review of pertinent literature, this paper delves into a comprehensive discussion from two perspectives: the integration of multiple datasets and the refinement of data features.

2.1.1. Combination of Databases and Clinical Test Sets

In the training of medical models, database information, owing to its advantages, including vast quantities, accessibility, and diverse disease representations, significantly facilitates data acquisition in the medical field, serving as the primary information source for model training. Nevertheless, the intricate nature of medical models necessitates that sole reliance on database information is insufficient to meet clinical requirements. Overly idealized data may induce overfitting during model training, thereby diminishing the model's real-world performance. Consequently, this paper advocates for the concurrent utilization

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

of database information and clinical test datasets to enhance model performance and explore its generalizability in clinical practice.

For instance, in the development of an artificial intelligence model for distinguishing benign from malignant pleural effusion, Sufei Wang et al. incorporated not only database information but also an external, independent dataset comprising 362 pleural effusion patients from Wuhan University Renmin Hospital to validate the model's efficacy **?**. In breast cancer management, the status of sentinel lymph nodes is paramount. Researchers achieved notable performance and generalizability by employing the ResNet 50 model, which was jointly trained using the TCGA database and clinical datasets from two hospitals **?**. Sebastian R. van der Voort et al. gathered data from five publicly accessible datasets to train a glioma classification model, further validating it with internal datasets from four distinct institutions, thus augmenting the model's generalization capabilities **?**. In their study of predicting the invasiveness of prostate cancer, Marija Pizurica et al. developed the TiDo model, which leveraged data from prostate cancer patients in the TCGA database and incorporated the internal dataset from UZ Ghent for collaborative training. This approach significantly enhanced the model's predictive accuracy **?**.

Furthermore, in non-pathological image domains, such as endoscopy, dual data training models have also exhibited promising performance. For instance, in training a model for polyp detection using colonoscopy data, researchers employed a fusion of five distinct datasets, encompassing nine colonoscopy videos. This integration of multiple data sources contributed to the improved generalizability of the model **?**. Xi Wang et al. initially conducted a comprehensive analysis of the SUCC dataset for lung cancer whole slide images (WSI) when applying a weakly supervised learning algorithm for lung cancer classification and diagnosis. Subsequently, they evaluated this methodology on the TCGA database. This cross-dataset training approach not only elevated the model's performance but also bolstered its generalizability in practical applications, thereby aiding clinicians in devising more precise and individualized treatment plans **?**.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

2.1.2. Selection of Data with Quality Features

The quality of data features is paramount in determining the efficacy of a trained model. In medical model training, challenges such as missing, erroneous, or ambiguous data features frequently emerge, posing significant threats to both clinical efficiency and model generalizability. Faced with these problems, this paper integrates and summarizes the existing literature, and analyzes the current solutions.

For instance, in the training of esophageal cancer classification models, Yoshimasa Horie et al. not only incorporated images representing various cancer types, including squamous cell carcinoma, adenocarcinoma, superficial cancer lesions, and advanced cancer lesions, but also underwent a rigorous preprocessing step to eliminate images with suboptimal feature representation due to artifacts like halos, blurring, defocusing, mucus, and inadequate air insufflation **?**. Similarly, in models aimed at gastric cancer molecular subtype classification **?**, breast cancer diagnosis **??**, and the prediction of colorectal cancer lymph node status **?**, researchers have prioritized the selection of detailed and comprehensive data. Some of these models have furthermore implemented a data preprocessing procedure, retaining only images with superior feature representation to mitigate potential model errors.

The integrity of data information and the representation of high-quality features serve as essential prerequisites for the successful training of models, and they are pivotal to their successful implementation in clinical practice. herefore, this paper will dive into the specification of data collection and summarize existing model methods to develop more effective models of clinical medicine.

2.2. Data Preprocessing

The training of deep learning models inherently depends on precise data annotations. However, clinical annotations often fall short of ideal standards due to practical constraints. As a result, the data preprocessing step assumes a pivotal role in model development. Traditional models frequently adopt the approach of segmenting images into smaller patches during preprocessing to

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

capture data feature information more accurately. While this technique enhances model performance, it concomitantly elevates the complexity of model computation.

To facilitate the broader application of deep learning models in the medical field, it is imperative to streamline data preprocessing steps and optimize the model computation process to the utmost extent, while maintaining model efficacy. This paper thus summarizes the following strategies to achieve this goal:

2.2.1. Semi-supervised Learning

Traditional machine learning classification methods primarily encompass supervised learning and unsupervised learning. Semi-supervised learning, on the other hand, marries the benefits of both approaches. By harnessing both labeled and unlabeled data, semi-supervised learning not only elevates the accuracy of machine learning models compared to conventional methods, but it also mitigates costs and enhances efficiency **?**. For instance, in a 2020 study conducted by Fernando P. dos Santos et al., it was elucidated that employing semi-supervised learning for data feature extraction can streamline model complexity while maintaining satisfactory performance **?**.

2.2.2. Supervised Learning

Supervised learning, a pivotal training paradigm in machine learning, employs labeled datasets to instruct algorithms for precise data classification and result prediction. Based on the nature of data annotation, supervised learning is further differentiated into strongly, weakly, and self-supervised learning. Especially, weakly supervised learning effectively trains models despite incomplete data labels, meeting expected performance standards. For example, in constructing disease prognosis models using brain MRI images, weakly supervised learning methods not only circumvent the limitation of limited data but also mitigate model complexity **?**. In models forecasting postoperative outcomes for patients with lung metastasis from colorectal cancer, researchers have also

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

leveraged weakly supervised learning to train convolutional neural networks for histological feature recognition, simplifying the process and attaining reliable prediction outcomes **?**. Additionally, self-supervised learning methods **?** capitalize on vast amounts of unlabeled data for training. For instance, in an eye disease detection model, the RETFound model proposed by Yukun Zhou et al. harnesses self-supervised learning on unlabeled retinal images, efficiently accomplishing the disease detection task **?**. The integration of supervised learning in model development ensures model validity while addressing excessive complexity, simplifying algorithmic procedures, and significantly contributing to the proliferation of deep learning models in the medical field.

2.2.3. Multi-instance Learning

In the realm of machine learning, multi-instance learning, an evolutionary approach stemming from supervised learning algorithms, exhibits significant potential for applications in medical. This potential stems from its fundamental concept of conceptualizing a collection of multiple instances as a "bag" **?**. Contrary to traditional supervised learning, which solely processes individually labeled instances, multi-instance learning engages with sets of labeled bags, each containing multiple instances. When addressing multi-instance binary classification problems, a cardinal principle is adhered to: if all instances within a bag are negative, the entire bag is designated as negative; conversely, if a single instance within a bag is positive, the bag is classified as positive. In the development of a liver cancer classification model, Chunli Sun et al. innovatively integrated multi-instance learning with global labels **?**. This approach not only achieved the anticipated model performance but also considerably streamlined data preprocessing procedures, thereby offering a novel perspective for model advancement.

2.2.4. Active Learning

Active learning **?**is an efficient machine learning technique aimed at optimizing model performance through the active selection of the most informative

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

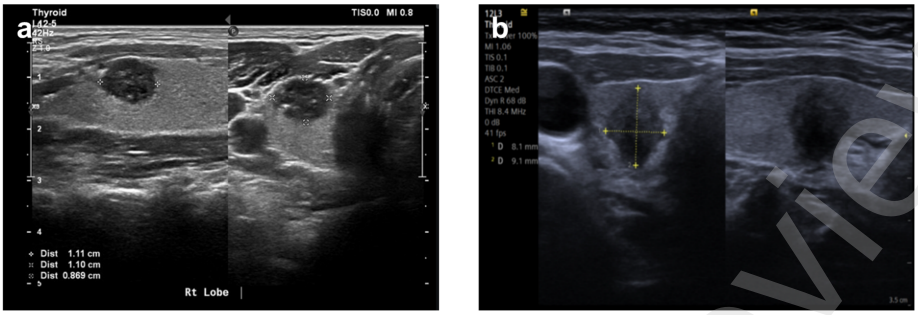

Figure 4: a. Ultrasound image of thyroid carcinoma b. Ultrasound image of parathyroid adenoma

samples for annotation, thereby utilizing minimal yet high-quality labeled data. Active learning serves as the cornerstone of the "human-in-the-loop" paradigm, and the integration of the two results in a synergistic human-machine collaboration model. This collaborative methodology not only minimizes the data preprocessing effort but also validates the reliability of the data, ultimately leading to enhanced model accuracy. In a 2021 study, Samuel Budd et al. explored the viability of active learning and provided an in-depth analysis of its potential applications, thereby contributing significantly to the simplification of the model's algorithmic workflow **?**.

This paper comprehensively outlines three distinct algorithmic strategies that can varyingly simplify data preprocessing steps and alleviate the complexity of the machine learning model's workflow, laying a robust foundation for its advancement in medicine and other domains. Furthermore, the introduction of automated lesion feature recognition modules **??** has emerged as a novel research focus in addressing pertinent challenges in this field, garnering significant attention due to their exceptional performance in reducing workflow complexity and opening new avenues for model development.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

2.3. Data Fusion

Given the intricate nature of medical diagnosis, healthcare professionals are often tasked with integrating and analyzing multi-dimensional data in clinical decision-making, which may necessitate the organization of multi-departmental consultations, thus posing novel challenges to clinical practice. To address these challenges and facilitate clinical practice, deep learning models necessitate innovations that transcend traditional data training paradigms, particularly in the exploration of novel data fusion methodologies. This article aims to provide a concise overview of data fusion training methods and achieved outcomes, laying a firm foundation for its progression in the medical domain.

2.3.1. Two Types of Data Fusion Methods

In a study conducted by Assaf Hoogi et al. in 2017, it was noted that due to the substantial differences in spatial lesion characteristics between CT and MR image sets, CNN models can be trained independently on these two distinct datasets **?**. This research further substantiated the efficacy of these models through training on a liver lesion dataset comprising 164 MRI and 112 CT images, highlighting the method's effectiveness, particularly in the classification of low-contrast and heterogeneous lesions. Furthermore, Claes Nøhr Ladefoged et al. introduced the concept of data fusion in their predictive model for brain dementia cohorts, utilizing both PET and MRI brain images to train the model, thereby enhancing its performance **?**. In 2022, Nico Sollmann et al. utilized a combined approach of vertebral CT and MRI images to train a model aimed at distinguishing osteoporotic vertebral fractures from non-osteoporotic fractures, aiming to enhance diagnostic efficiency, and achieved promising diagnostic outcomes **?**.

2.3.2. Multi-type Data Fusion Method

Aurora Rosvoll Groendahl et al. pioneeringly employed CT, PET, and MRI images to collaboratively train a CNN model in their approach to address the challenge of target delineation for anal squamous cell carcinoma (ASCC) **?** (Fig-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

ure 4). The experimental outcomes indicate that this model is capable of delivering high-quality automatic target delineation for both single-modal and multi-modal image inputs, thereby indicating the potential of deep learning as a universal instrument for target volume delineation in ASCC patients in the future. Furthermore, in the realm of ischemic stroke research, the ischemic core and defect volume serve as pivotal imaging biomarkers that hold a significant influence on patient treatment decisions. To address this, Chengyan Wang et al. introduced a multi-scale 3D convolutional neural network as a prediction model in their 2021 study, utilizing normal CT, CTA, and 8-second delayed CTA images for joint training, thereby aiding in the diagnosis and treatment of clinical conditions **?**.

The advent of data fusion has indisputably accelerated the advancement of deep learning in the medical domain, resulting in a substantial enhancement in the efficacy of disease diagnosis. Nonetheless, current data fusion techniques primarily concentrate on imaging-based tasks, thereby leaving a void in the realm of pathology. The integration of imaging manifestations with histological features holds paramount importance in elevating the precision of pathological diagnosis. Consequently, future research ought to delve further into the specific methodologies and applications pertaining to this area.

3. Model

In the application of deep learning models, interpretability, robustness, and generalization pose significant challenges. Specifically, the lack of interpretability impedes clinical staff's full trust in AI-assisted applications. Simultaneously, deficiencies in robustness and generalization constrain the stability and universal applicability of these models. Collectively, these factors hinder the widespread adoption of deep learning models in the medical field. Therefore, to enhance the practicality of model applications and foster the robust development of artificial intelligence in healthcare, this paper investigates strategies to mitigate these limitations, aiming to transform the current state of affairs.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

3.1. Interpretability

The issue of interpretability in deep learning algorithms constitutes a significant obstacle to their widespread adoption in clinical medicine. This impediment primarily stems from the absence of a fundamental breakthrough in the "black box theory" **?**. As deep learning algorithms frequently operate in a "black box" fashion, their output results are often inscrutable and difficult to justify for human comprehension. This lack of interpretability is particularly acute in medical , where human safety is paramount, thus fostering distrust between doctors and patients. This distrust not only constrains the promotion and developmental prospects of artificial intelligence but also undermines the general acceptance of these models, creating a detrimental vicious cycle.

To mitigate this challenge, this paper delves into various approaches to enhance the interpretability of models from diverse perspectives, thereby laying the groundwork for their advancement in medicine and beyond.

3.1.1. Model Interpretability

During the diagnostic procedure, accurate localization of pathological alterations is pivotal. Nevertheless, traditional approaches fall short in providing sufficient elucidation regarding the precise delineation of lesion areas by the model. Consequently, enhancing the interpretability of the model's lesion identification process is of utmost importance. As an illustration, by integrating image binarization with the Otsu thresholding method, histological images can be transformed into grayscale representations. Subsequently, the lesion regions within these grayscale images can be delineated using watershed segmentation techniques **?**. This approach provides an intuitive understanding of the lesion recognition process of the model, thereby alleviating the inherent abstraction of machine learning algorithms and fostering greater trust between healthcare professionals and patients. In addition, the use of CAM (class activation mapping) images facilitates the visualization of deep learning features, which in turn produces heatmaps closely consistent with histological features, further improving the interpretability of the model reference **?**. Pierre Courtiol et al implemented a

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

scoring system in MesoNet to quantify histological features associated with prognosis citation to facilitate the selection of further clinical analysis and utilized critical data based on the scoring results **?**.Additionally, the binary classification method employed by William Lotter et al. **?**offers a degree of interpretability to the model's precise lesion identification process, thereby alleviating varying degrees of trust concerns among clinicians.

3.1.2. Interpretability of Results

Explaining the operational patterns of the model from a classification perspective serves to mitigate trust concerns among clinicians and patients towards deep learning algorithms. This paper comprehensively outlines three approaches to achieve this: Firstly, visualizing classification outcomes through the correlation between images and diagnostic results facilitates the intuitive depiction of diagnostic evidence. For instance, in a glioma classification model, the visualization process correlates classification results with histological glioma subtypes, thereby enhancing the credibility of diagnostic outcomes **?**. Secondly, the construction of knowledge-driven modules and utilization of pertinent theories for visually elucidating the model's classification process aids in addressing interpretability challenges of traditional models **?**. Finally, clinical research exploring the influence of diverse data on the interpretability of classification results, such as leveraging asymmetric pixel values in MRI scans of the brain to enhance feature interpretability in early dementia diagnosis models, further advances this field **?**.

This paper extensively examines strategies to enhance algorithm interpretability from both model and classification perspectives, aiming to alleviate trust issues among clinical personnel and establish a foundation for the widespread adoption of deep learning algorithms. However, it is imperative to note that elucidating the model through theoretical transformations or visual analyses may augment architectural complexity and necessitate clinicians to possess a certain degree of theoretical proficiency. Consequently, further exploration is warranted in simplifying the model architecture to enhance its comprehensibility and ac-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

ceptance among clinical staff in addressing interpretability challenges.

3.2. Robustness

"Robustness" pertains to the inherent capability of a control system to sustain stability in specific performance metrics despite structural or magnitude alterations of its parameters within a predefined range. In the realm of deep learning, robustness is predominantly employed as a metric to assess the stability of a model under diverse conditions, particularly in the presence of data fluctuations and external disturbances, determining whether the model can sustain consistent performance. The significance of robustness is paramount for the practical deployment of models, particularly in the medical field. Given the intricate and variable nature of clinical scenarios, the stability of the model is directly correlated with the ultimate decisions made by patients and medical professionals.

To address the issue of robustness and facilitate the widespread adoption of models in the medical domain, this paper comprehensively outlines the following key considerations:

3.2.1. Consideration from a Data Perspective

Traditional models frequently rely solely on annotations from a single expert or data from a single platform, thereby inherently restricting their robustness and compromising their ability to accurately handle intricate data inputs. Hence, this paper underscores the importance of adopting multi-expert, multi-platform annotation methodologies. This approach serves to augment the informational content of the training data and bolster the model's capacity to process diverse input data. For instance, Satoshi Fujii and colleagues capitalized on data gathered from multiple centers to develop a colorectal cancer gene screening model, significantly bolstering its robustness **?**. Analogously, Yinxi Wang and her team incorporated data from two platforms in their breast cancer prediction model for intratumoral heterogeneity, resulting in both improved performance and a marked enhancement in the model's robustness **?**.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

Furthermore, in this paper, we have also employed multi-center, multi-platform data sources in our models for lymph node melanoma metastasis **?** and the recognition of tumor-infiltrating lymphocytes in primary melanoma **?**, thereby providing robust evidence for the enhanced robustness of our models.

In developing a model for detecting microsatellite instability in early colorectal cancer, Klaus Gerwert et al. procured clinical data and tissue samples from 188 early colorectal cancer (CC) patients spanning Germany **?**, thereby ensuring the robustness of the model. In 2017, Harshita Sharma et al. employed detailed annotations from ten pathologists to train a deep learning model for gastric cancer identification **?**. Similarly, Swiderska-Chadaj et al. highlighted in their 2019 article that the data information was thoroughly annotated by three pathologists **?**, all of which contributed significantly to the model's robustness and enhanced applicability in clinical practice. The utilization of multi-center, multi-expert annotated data incontrovertibly enriches the data content of the model, enhances its stability in complex scenarios, and positively contributes to the overall robustness of the modeling framework.

3.2.2. Consideration from a Model Perspective

In the realm of model design, researchers are consistently pursuing novel methodologies to bolster robustness. For instance, in the 2023 study conducted by Yufei Zhou et al., they presented an innovative Multi-Site Cross-Organ Calibration (MuSClD) framework grounded in deep learning for the automated diagnosis of non-melanoma skin cancer **?**. By calibrating whole slide images (WSI) from non-target organs, obtained at the same site as the target organ, they achieved a noteworthy enhancement in the model's robustness. Furthermore, in their tumor heterogeneity diagnosis model, the authors introduced a Multi-Instance Learning with Multiple Detection (MDMIL) methodology **?**, which significantly improved the robustness and diagnostic capabilities of the model through the utilization of multiple detection strategies and the imposition of feature constraints between samples. Additionally, Varun Gulshan et al. developed an algorithm with adjustable operating points for their deep learning

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

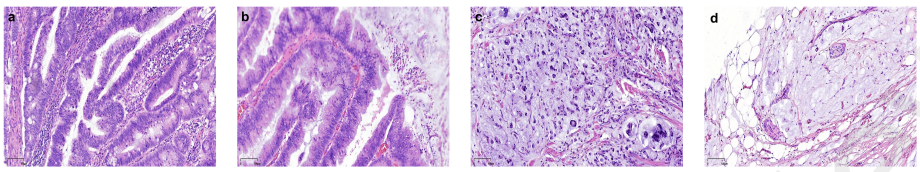

Figure 5: Pathological images of colorectal cancer: a.Adenocarcinoma, with distorted glands and fused glandular lumens forming sieve-like structures visible. b.Adenocarcinoma with mucinous secretion, featuring a significant amount of mucinous secretion visible in the upper right corner on top of the adenocarcinoma background. c.Signet-ring cell carcinoma, characterized by numerous signet-ring cells floating above the mucus, with nuclei displaced to one side, resembling a ring. d.Adenocarcinoma invading nerves, with numerous tumor cells and mucus visible around the nerves.

model for diabetic retinopathy detection in fundus photographs **?**. This flexibility allows for the optimization of sensitivity and specificity based on clinical requirements, thereby enhancing the practicality and robustness of the model.

This paper provides an in-depth exploration of strategies to address the robustness limitations of deep learning algorithms, encompassing both data and model-centric perspectives. However, it is crucial to acknowledge that in medical image data analysis, minute variations in data originating from diverse devices may contribute to diagnostic inaccuracies, thereby jeopardizing the precision of clinical decision-making. While studies, such as the cirrhosis portal hypertension detection model proposed by Yanna Liu et al., claim that their model demonstrates consistent detection performance regardless of the imaging data source machine **?**, the underlying mechanisms and operative procedures necessitate further scrutiny and enhancement. Consequently, to strengthen the robustness of deep learning models in future research, it is imperative to prioritize data standardization and enhance model generalization capabilities, thereby facilitating the broader application of deep learning in the medical domain.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

3.3. Generalization

In traditional deep learning models, the complexity of disease mechanisms and the diversity of histological features often hinder the models' accurate diagnosis of similar diseases. This challenge is primarily attributed to the impairment of the models' generalization performance, which pertains to the adaptability of machine learning algorithms to novel samples. Specifically, generalization performance measures whether an algorithm can provide precise outputs for data that falls outside the learning set but retains the same underlying patterns. The presence of these issues has exacerbated the diagnostic performance fluctuations in existing models, thus impeding the effective application of machine learning methods in clinical practice and the satisfaction of clinicians' actual needs. To tackle this obstacle, this paper delineates the following strategies to enhance generalization performance:

3.3.1. Optimizing Model Structure

Despite the extensive history of the integration of machine learning and medical imaging, traditional convolutional neural network (CNN)-based algorithms often encounter challenges in maintaining consistent accuracy when confronted with extensive and intricate datasets coupled with variable clinical characteristics. This predicament can potentially culminate in diagnostic inaccuracies, thereby influencing clinical decision-making. To mitigate this issue, researchers have introduced the paradigm of transfer learning, which aims to augment the model's capacity for generalization in handling complex data **???**. For instance, the GoogleNet-based transfer learning algorithm, leveraging pre-trained models for glomerular data training, not only expedites the training process but also substantially elevates the model's generalization performance, significantly enhancing its performance in clinical applications **?**.

3.3.2. Utilizing Multi-center Data

Data serves as the cornerstone of machine learning, and the abundance and quality of clinical data are crucial factors influencing the efficacy of model train-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

ing and performance in real-world applications. Presently, data employed in the realm of medical imaging and deep learning is predominantly derived from publicly available repositories, such as TCGA, and clinical datasets from select hospitals. However, the standardized nature of public databases may constrain the model's applicability in diverse clinical settings. Additionally, clinical data is often underutilized due to variations in geographical regions, ethnicities, and other confounding factors, potentially compromising the model's generalization capabilities and universal applicability. To address these limitations, researchers have advocated for the employment of multi-center data **??**, which can encapsulate the unique characteristics of different regions and populations, mitigate data randomness, and minimize the impact of non-disease-related factors on model performance. For instance, Yinxi Wang et al. integrated patient data from two distinct sources in the training and internal validation of the Deep-ITH model. Their approach involved evaluating the model's performance across two external test sets **?**, resulting in significant enhancements in overall performance. The incorporation of multi-center data has further bolstered the model's generalization abilities, thereby laying a robust foundation for its clinical deployment **?**.

3.3.3. Introduction of Multi-expert Annotations

Deep learning has significantly expedited diagnostic procedures in the medical field, previously lengthy and time-consuming. However, the intricate nature and variability of medical images, coupled with confounding factors such as tumor heterogeneity and artifacts, pose challenges for non-experts in accurately interpreting image information. This often leads to errors in data information, thereby compromising the model's performance and generalization capabilities. To mitigate this issue, this study advocates for the involvement of multiple experts with pertinent expertise during the data preprocessing stage, to provide meticulous annotations and enhance data quality. This approach can subsequently facilitate the training of models, enabling them to achieve clinically acceptable levels of accuracy and generalization. For instance, Harshita Sharma

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

et al. **?**utilized annotations from 10 pathologists in 2017 to train a deep learning model for gastric cancer recognition, while Swiderska-Chadaj et al. **?**employed annotations from 3 pathologists in 2019 to train the U-Net model, both studies achieving commendable generalization performance and clinical applicability.

In conclusion, evaluating the generalization ability of models is pivotal in their clinical application, as it directly determines their efficacy in supporting clinicians and attaining broad applicability. By addressing the limitations of traditional medical models, this article outlines practical strategies for improvement, encompassing model structure, data source, and data processing procedures.

4. Output

Deep learning models exhibit extensive applications in the medical field, encompassing primarily segmentation **?**, diagnosis **?**, and classification **?**as their core tasks. These fundamental tasks further specialize into refined subtasks within various disease domains. Owing to the continuous advancement of deep learning technology, these tasks have achieved remarkable performance enhancements in the medical sector. The present paper endeavors to provide a concise overview of common models pertaining to these three tasks, while also delving into their current limitations and prospective avenues for future development.

4.1. Segmentation Task

As a pivotal application of deep learning within the medical domain, the segmentation task centrally concerns the precise identification of lesion areas in medical images, thereby serving as a cornerstone for subsequent diagnostic and classification procedures. Achieving accurate segmentation necessitates not only reliance on expert annotations detailing histological features but also the effective mitigation of artifacts, noise, and interferences originating from non-lesion regions. Consequently, the efficacy and precision of the segmentation process are paramount for accurate disease classification and diagnosis.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

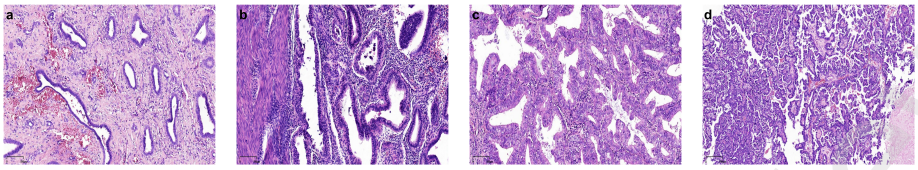

Figure 6: Pathological images of benign and malignant uterine lesions: a. Proliferative phase endometrium, characterized by dilated endometrial glands of varying sizes. b. Secretory phase endometrium, featuring distorted endometrial glands with visible secretions within the glandular lumens. c. Endometrioid adenocarcinoma with tubular structure, displaying fused glandular lumens forming sieve-like patterns, and increased nuclear pleomorphism visible in the cells. d. Endometrioid adenocarcinoma with papillary structure, exhibiting numerous papillary and micropapillary formations, with blood vessels traversing between the papillae. Among these, a and b represent benign lesions of the endometrium, while c and d are two distinct manifestations of endometrioid adenocarcinoma.

4.1.1. U-Net Model

The U-Net model **?** stands as one of the seminal and effective algorithms that harness fully convolutional networks for semantic segmentation. This model comprises two distinct yet interconnected parts: a contracting path, which encodes contextual information from the image, and an expanding path, which precisely locates the target regions. These two paths are symmetrically arranged, thereby forming a distinctive U-shaped architecture. The U-Net model has exhibited outstanding performance in various medical image segmentation tasks. For example, in the task of differentiating Bowen's disease from seborrheic keratosis, Philipp Jansen et al. trained a U-Net model to achieve precise segmentation of Bowen's disease, seborrheic keratosis, and normal tissues, demonstrating high sensitivity **?**. Furthermore, the U-Net model has been employed in the development of prognostic models for oral squamous cell carcinoma. By segmenting and quantifying the epithelial layer, it provides a crucial foundation for clinical assessment of cancer prognosis, thereby lessening the reliance of

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

pathologists on histological diagnosis **?**.

In the model proposed by Daniel Firmbach et al. in 2023 for predicting the tumor-stroma ratio in colorectal cancer, the U-Net model exhibited remarkable proficiency in segmenting tumor regions in hematoxylin and eosin (HE)-stained whole slide images (WSIs) of colon cancer patients into five distinct categories: tumor, stroma, necrosis, mucus, and background (Figure 5). Its accuracy significantly surpassed manual segmentation, thereby providing invaluable insights for clinical prognosis evaluation and subsequent therapeutic decisions **?**. Furthermore, in the realm of endometrial cancer screening, Qing Li et al. harnessed the U-Net model to automatically identify and diagnose pathological images of endometrial cell clusters (ECCs), efficiently discriminating between malignant and benign groups (Figure 6). Post-training, the model displayed remarkable specificity and sensitivity, bolstering the development of an artificial intelligence system for endometrial cancer diagnosis **?**. Similarly, in the context of COVID-19 detection and diagnosis, Rokaya Rehouma et al. corroborated the exceptional performance of the U-Net model in lung lesion segmentation **?**.

The continual evolution and innovation of the U-Net model have propelled significant advancements in medical segmentation tasks, considerably expanding the applicability of deep learning algorithms in the medical domain. Nevertheless, as medical data accumulates and task complexity increases, future research necessitates ongoing refinement and innovation in terms of model accuracy, generalization capabilities, and computational efficiency.

4.1.2. Exploration of Other Models

Although the U-Net framework, rooted in convolutional neural networks, has established itself as a benchmark for medical image segmentation, its performance remains imperfect. In a study conducted in 2020, Tao Wang et al. highlighted the limitations of the U-Net framework in capturing global and long-range semantic information **?**. Recognizing the prowess of the Transformer structure in capturing global information and the proficiency of convolutional neural networks (CNNs) in local information processing, this study introduces a

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

novel O-Net network structure that adeptly amalgamates the distinct characteristics of CNNs and Transformers. This novel structure aims to comprehensively integrate global and local information, thereby enhancing the performance of medical image segmentation and classification. Experimental outcomes validate that the O-Net model exhibits robust generalization capabilities and competitive performance in both segmentation and classification tasks, offering valuable insights for the advancement of U-Net-based models.

Deep learning serves as a pivotal technology in the medical domain, particularly in advancing histology-based nuclear segmentation and tumor microenvironment characterization, where the current research endeavors to delve into novel model architectures. Specifically, the proposed histology-based Yolo model markedly enhances nuclear segmentation speed while surpassing existing WSI analysis methods in nuclear detection, classification accuracy, and computational efficiency **?**. To address the challenge of excessive time consumption in segmenting cervical cell nuclei, particularly in large-scale cellular images, Die Luo et al. introduced an innovative dual-supervised sampling network model in 2023 **?**. This model efficiently reduces the computational load in image feature extraction via the implementation of a supervised downsampling module and a boundary detection network, all while upholding segmentation accuracy. Experimental findings indicate that, in comparison to U-Net, this model achieves a fivefold increase in inference speed while maintaining segmentation precision.

This article presents a comprehensive review of diverse approaches to enhance the precision of medical image segmentation tasks, encompassing multiple modeling perspectives. Nevertheless, current models still face certain challenges. For instance, Hiroyuki Abe et al. developed an artificial intelligence-driven gastric biopsy pathological diagnosis system employing a multi-stage pathological semantic segmentation (MSP) model in a 2022 study. Despite the system's demonstrated potential in routine clinical practice across various institutions, further in-depth exploration and research are imperative to elucidate the model's internal mechanisms and practical feasibility **?**. Future research endeavors should strive to deepen the comprehension of these models, thereby

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

facilitating further advancements in medical image segmentation.

4.2. Diagnostic Tasks

The accurate diagnosis of diseases holds a pivotal position in the medical field, profoundly influencing subsequent treatment planning and prognosis. Nonetheless, relying solely on clinicians' expertise and decision-making often falls short in comprehensively encompassing all diseases, leaving the results prone to subjectivity. However, the advent of deep learning technology has significantly enhanced the precision and objectivity of medical diagnosis. By constructing and training diagnostic models, coupled with manual refinement, deep learning has attained remarkable clinical efficacy in this domain. The present article endeavors to offer a concise overview of the prevailing diagnostic models.

Traditional diagnostic methods primarily harness Convolutional Neural Network (CNN) models. For example, Niha Beig, MS, et al. proposed a lung disease diagnostic model in 2019 that proficiently discriminated adenocarcinoma from granuloma by leveraging radiomic features on lung CT images **?**. Nonetheless, as research progresses, numerous models based on CNN have been innovated and optimized, including GoogleNet, AlexNet, ResNet, and VGG16.

4.2.1. GoogleNet

GoogleNet, originating from the ImageNet model, has undergone numerous advancements, resulting in the development of subsequent versions such as Inception V2, Inception V3, and Inception V4. These iterations have effectively tackled challenges posed by traditional models, including excessive parameters, high computational complexity, and gradient dispersion. Moreover, they have gained significant traction in the realm of medical diagnosis. For instance, in 2022, Jianglin Zhang et al. opted for the GoogleNet model in developing a diagnostic system for diabetic foot ulcers. Through rigorous multi-angle comparisons, they corroborated the model's superior classification accuracy compared to alternative models **?**. Similarly, Shruti Kannan et al. employed a pre-trained

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

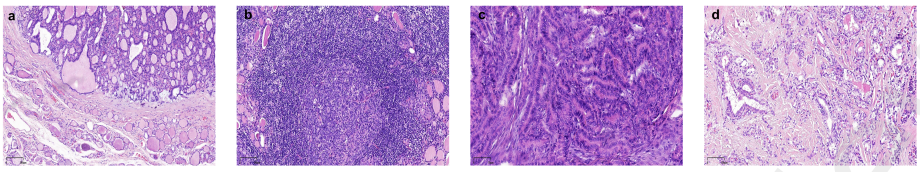

Figure 7: Pathological images of benign and malignant thyroid lesions: a. Thyroid adenoma, with a distinct fibrous capsule visible between the lesion and normal tissue in the image. b. Lymphocytic thyroiditis, characterized by focal lymphocyte aggregations around the thyroid follicles. c. Classical subtype of thyroid carcinoma, featuring loss of normal thyroid follicular architecture, increased cellular pleomorphism, ground-glass appearance of nuclei, and the presence of inclusions, nuclear grooves, and other such structures. d. Follicular subtype of thyroid carcinoma, where the cancerous tissue is entirely follicular in structure, devoid of papillary formations, and exhibits increased cellular pleomorphism. Among these, a and b represent benign thyroid lesions, while c and d are two different common subtypes of thyroid cancer.

GoogleNet model for glomerular data training in the context of glomerulosclerosis diagnosis, yielding promising diagnostic outcomes **?**.

4.2.2. AlexNet

AlexNet, a classical convolutional neural network introduced during the ImageNet image classification competition, significantly surpasses traditional neural networks in terms of depth, allowing it to learn the intricacies of datasets with greater efficacy. Furthermore, AlexNet incorporates data augmentation and Dropout modules, which serve to mitigate overfitting and enhance the precision of image classification. In a recent 2023 study conducted by Huayi Zhao et al., the AlexNet model was employed to distinguish Graves' disease, subacute thyroiditis, and thyroid tumors (Figure 7), demonstrating high levels of sensitivity and specificity **?**.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

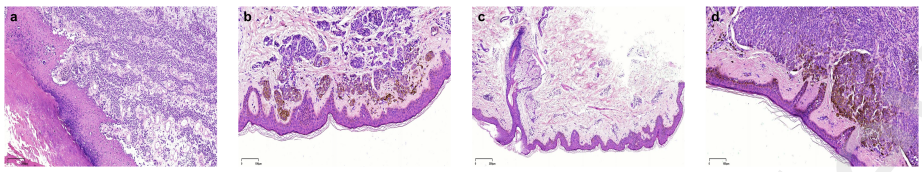

Figure 8: Pathological images of benign and malignant skin lesions: a. Melanoma, characterized by an infiltrative distribution of a large number of tumor cells beneath the epidermis. b. Intradermal nevus, with nevus cells localized within the dermis. c. Junctional nevus, featuring nevus cells distributed within the basal layer of the epidermis. d. Compound nevus, where nevus cells are distributed in both the epidermis and the dermis. Among these, a represents a malignant skin lesion, while b, c, and d are benign skin lesions.

4.2.3. ResNet

The fundamental component of the ResNet model resides in its residual unit design, facilitating direct conveyance of inputs via cross-layer connections and executing identity mapping. Subsequently, the residual unit integrates the result of this mapping with the outcome of convolutional transformations. This distinctive architecture significantly alleviates the challenge of gradient vanishing in deep neural networks. In the realm of skin cancer diagnostic models, Roman C. Maron et al. attained favorable diagnostic outcomes using the ResNet model, especially in the identification of skin melanomas and nevi (Figure 8). Additionally, the ResNet model exhibited outstanding performance in diagnosing thyroid cancer lymph node metastasis **?**. A recent study conducted in 2023 by Yang Guang et al. highlighted that the utilization of the ResNet model for processing multimodal data from PTC patients can produce precise diagnostic results, thereby enhancing clinical decision-making and subsequent treatment strategies **?**.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

4.2.4. VGG16

The VGG model exhibited remarkable proficiency in the 2014 ImageNet image classification competition. Especially, VGG16 and VGG19 garnered significant attention owing to their superior performance. These models leverage consecutive 3x3 small convolution kernels and pooling layers to construct deep neural networks, achieving a network depth of either 16 or 19 layers. In the context of IgA nephropathy diagnosis, Ying Chen et al. introduced a diagnostic model grounded in VGG16, which accurately interprets the immunofluorescence outcomes of IgA nephropathy, surpassing the coincidence rate of pathological diagnosis by medical experts **?**. Additionally, VGG16 has been implemented in a diagnostic model for papillary thyroid carcinoma. A 2021 study by Yi-Jia Lin et al. revealed that this model's application enhances diagnostic efficiency while minimizing the impact of subjective factors, exhibiting exceptional accuracy **?**. The integration of the VGG16 model with pathological diagnosis has the potential to optimize pathological workflows, bolster the efficiency of pathologists, and mitigate misdiagnosis rates stemming from reading fatigue, thereby rendering pathological diagnosis more precise and impartial.

In conclusion, the advent of diagnostic models has provided clinicians with valuable assistance, effectively addressing persistent clinical challenges. However, to achieve widespread clinical application, further refinement in model accuracy is imperative. By integrating diverse data sources, incorporating clinical expertise, and continually refining the model architecture, we anticipate providing enhanced services for clinical practice.

4.3. Classification Tasks

In the field of medicine, precise disease classification is paramount for devising tailored treatment plans, surgical approaches, and medication regimens. However, the classification of specific diseases is often prone to errors stemming solely from clinicians' experience and subjective judgment. Consequently, the incorporation of deep learning algorithms to aid in clinical classification has emerged as a pivotal research focus with profound practical implications. This

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

article endeavors to critically evaluate and delve into existing deep learning classification models.

Convolutional neural networks (CNNs) currently constitute the primary choice for classification tasks. By advancing or innovating CNNs to varying degrees, the accuracy of classification tasks can be further optimized, thereby aiding clinicians in making informed clinical decisions and treatment selections. Furthermore, transfer learning has garnered widespread application in medical classification tasks and has demonstrated commendable efficiency outcomes.

4.3.1. Data-rich Disease Classification

The DenseNet model, a convolutional neural network variant, demonstrates superior performance compared to traditional convolutional neural networks, achieving this with fewer parameters and reduced computational costs, attributed to the dense connections it establishes between all preceding and subsequent layers. In a 2022 study, Nadine Flinner et al. proposed that the DenseNet model can accurately identify four molecular subtypes of gastric cancer, significantly outperforming immunohistochemical experimental results **?**. For the classification of glioma subtypes, traditional invasive genetic testing has been the conventional approach. However, in a 2021 study, Julia Cluceru et al. developed a CNN-based classification model utilizing VGG and ResNet models, effectively classifying genetic subtypes in a non-invasive manner. This approach is more conducive to clinical promotion and adoption **?**. In the context of blood disease diagnosis, Mazen Osman et al. employed various CNN networks to categorize monocytes, promonocytes, and monoblasts in blood samples. Their findings provide a crucial foundation for leukemia subtype classification, thereby guiding clinical treatment strategies **?**.

Moreover, in a 2021 study by Mazen Osman et al., a thyroid papillary carcinoma gene classification model, utilizing the Inception V3 architecture, was formulated. This model attained a diagnostic accuracy of 95.2% for BRAF and RAS gene mutations on an independent testing subset, thereby showcasing the efficacy of deep learning techniques in aiding pathologists in cancer histo-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

logical classification**?**. In the subcategorization of osteosarcoma, pathologists often encounter difficulties stemming from histological morphological compression, inter-class similarities, intra-class variations, and noise. However, in a 2023 study conducted by Md. Tarek Aziz et al., through the integration of a CNN with a multilayer perceptron (MLP) model and training on a whole slide image (WSI) dataset, they effectively classified three osteosarcoma subtypes, significantly surpassing existing methods and offering robust support to clinicians in osteosarcoma diagnosis **?**. In classifying non-Hodgkin's lymphoma, Georg Steinbuss et al. employed CNN models to categorize data from 629 patients, thus providing valuable supplementary information for clinical pathologists **?**. Furthermore, in a lung cancer subtype classification model, Mark Kriegsmann et al. utilized the InceptionV3 CNN model and achieved promising results, offering significant support for precision treatment of lung cancer **?**.

4.3.2. Disease Classification with Limited Data

For diseases with limited data availability, traditional convolutional neural networks often encounter difficulties in supporting model training. In addressing this challenge, transfer learning methods have garnered significant attention. Yang Zhang et al. implemented a two-center-based transfer learning model for breast cancer molecular subtype classification, achieving a substantial improvement of over 30% in classification accuracy through parameter refinement. This demonstrates the efficacy of transfer learning in mitigating data insufficiency **?**. Similarly, in the classification of functional and non-functional adenomas, Hongyu Li et al. adopted a transfer learning algorithm, leveraging a pre-trained segmentation model to extract features from MRI images. Their approach achieved high-performance results across multiple datasets, further affirming the advantages of transfer learning in medical classification tasks **?**. Furthermore, transfer learning has been employed for model enhancement. For instance, in a 2023 study, Riaz Ahmad et al. utilized transfer learning to augment a previous white blood cell classification model, achieving remarkably high accuracy, thereby providing a robust tool for leukemia diagnosis**?**.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

In summary, the application of deep learning classification models in the medical domain, encompassing both solid and non-solid tumors including leukemia, has gained increasing prevalence. These models have significantly contributed to clinical diagnosis, prognosis, and the optimization of treatment plans. However, despite these advancements, existing models still exhibit limitations in multi-class disease classification. Future research ought to delve deeper into techniques that integrate data fusion with model optimization, thereby advancing the accuracy of model classification and furthering the development of the medical field.

5. Future Development Directions

Utilizing the "input-model-output" framework, this review comprehensively evaluates the current state and viability of deep learning networks in medical imaging. For input date, challenges such as Data volume, data processing process, and data fusion are systematically elaborated upon, providing a comprehensive and structured understanding of these concepts. For AI models, we discuss limitations pertaining to generalization capacity, robustness, and interpretability, alongside strategies for improvement. Additionally, we delve into the application progress and current research standing of existing models for medical applications encompassing segmentation, diagnosis, and classification. Nonetheless, it is noteworthy that this discussion does not comprehensively encompass all challenges encountered by deep learning in the medical domain, thereby necessitating further detailed deliberations in various aspects.

5.1. Integration of Interdisciplinary Expert Knowledge

The intricate pathogenic mechanisms of diseases manifest in diverse clinical presentations, with even nuanced variations among subtypes of the same disease, thereby posing significant challenges for accurate diagnosis. To achieve refined diagnostic outcomes, relying solely on a single data source for model training is insufficient for current disease diagnosis models. While some research has focused on intra-disciplinary data fusion, such as integrating CT, MRI, and PET

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

data within the imaging domain, which has moderately enhanced model diagnostic performance, its fusion impact remains limited and falls short of fulfilling clinical demands. Conversely, inter-disciplinary data fusion harbors greater potential. For instance, integrating histological data with imaging modalities like CT and MRI can further augment the diagnostic prowess of the model, offering invaluable assistance for clinical practice **?**. However, the advancement of inter-disciplinary data fusion is still nascent, lacking robust literature support and clinical validation. Furthermore, it necessitates collaborative efforts from experts across multiple disciplines for precise data annotation.

5.2. Development of Novel Medical Models

In contemporary medical tasks, deep convolutional neural networks prevail as processing models, yet their diagnostic efficacy exhibits variability when confronted with nuanced lesions, faintly stained pathological images, and imaging artifacts. To address clinical demands, the development of novel models tailored for medical fusion applications has emerged as a focal research area. Through the refinement of processes, the augmentation and refinement of existing models, and the integration of knowledge-driven modules, not only can the diagnostic accuracy of these models be enhanced, but their interpretability can also be improved **??**. The development of knowledge-driven modules necessitates the involvement of experts across various fields, leveraging their professional knowledge for precise disease annotation. This approach not only aids in addressing the diagnostic challenges posed by rare diseases, but also elevates the interpretability of the models, thereby fostering their widespread clinical adoption. The introduction of these novel models enriches the application landscape of deep learning in the medical domain, marking a significant milestone; however, further validation of their feasibility through comprehensive databases and clinical datasets remains imperative.

5.3. Construction of Non-Solid Tumor Models

Although deep learning has been extensively employed in medical tasks, including disease segmentation, diagnosis, and classification, current research

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

predominantly focuses on solid tumors, such as liver cancer **?** and lung cancer **?**, with limited attention to non-solid tumors. However, as the incidence of non-solid tumors, such as leukemia, continues to rise annually, the importance of early diagnosis and classification for optimal treatment selection and patient prognosis has become paramount. Therefore, employing deep learning models to address challenges in disease diagnosis and classification of non-solid tumor medical images presents substantial research opportunities. For instance, utilizing deep learning models for identifying and annotating blood cell images can rapidly and precisely facilitate diagnostic and classification outcomes for non-solid tumors like leukemia, thereby assisting clinicians in devising effective treatment plans and improving patient outcomes **??**. Although current research has yet to comprehensively address the application of deep learning in non-solid tumors and lacks extensive data support, its potential clinical significance remains significant, indicating promising applications.

5.4. Integrated Multi-Data Website Platform

The current utilization of deep learning models in medical applications is limited, primarily serving diagnostic or classification functions. However, contemporary clinical practice demands a more comprehensive understanding and multifaceted analysis of diagnostic outcomes. For instance, in the field of pathology, existing machine learning models narrowly focus on histopathology, overlooking the pivotal complementary roles of molecular and cytopathology in clinical practice. To address this inherent bias, the development of an integrated multi-data web platform is paramount. This platform should incorporate conventional histopathological data, encompassing gross tissue information, histological details, and pathological alterations. Furthermore, it should encompass molecular pathological data (e.g.,BRAF gene mutation status in thyroid cancer **?**), cytopathological data (e.g. pleural effusion **?** and ascites **?** in lung and liver cancer patients), as well as data related to pathological immunohistochemical staining **?**and specialized stains **??**. This comprehensive data not only offers specific insights for pathological change classification and differentiation, but also

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

facilitates lesion staging and grading, thus significantly guiding surgical interventions, targeted drug therapies, and post-treatment care. The establishment of such an integrated data platform mitigates diagnostic errors in deep learning models stemming from factors like race, gender, and age, while minimizing accidental errors through the integration of diverse data sources. Nevertheless, the successful implementation and utilization of this platform necessitate the collaborative efforts of a wider range of pathologists.

6. conclusion

This paper conducts a comprehensive analysis of the current status of deep learning models' applications in the medical field, delving into the shortcomings of existing models from an "input-model-output" perspective. The paper addresses pivotal issues in the medical domain, including data quality, model architecture design, and application scope, and proposes refinement strategies while summarizing effective methodologies. In contrast to traditional reviews, this paper underscores the integration of models with clinical practice. With a focus on enhancing the practical utilization of models, it critically examines the challenges encountered by deep learning models in clinical settings, aiming to bridge the divide between theory and application.

Moreover, this paper highlights understudied areas such as the integration of interdisciplinary expert knowledge, the development of novel medical models, the evolution of non-solid tumor models, and the establishment of integrated multi-data web platforms. The introduction of these potential research avenues holds substantial clinical and research implications, offering novel exploration pathways for future researchers and contributing to addressing the current gaps in integrated development, thereby fostering the continuous advancement of deep learning in the medical domain.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167

Compliance with Ethical Standards

Declaration of competing interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments

This work is partially supported in part by the National Natural Science Foundation of China under grants Nos. 61976102 and U19A2065; the International Cooperation Project under grant No. 20220402009GH; the National Key R&D Program of China under grants Nos. 2021ZD0112501 and 2021ZD0112502;Natural Science Foundation of Jilin province(20210101252JC)

Credit authorship contribution statement

Bing Liu: Conceptualization of this study, Methodology, Software, Writing - original draft.
Yujia Cong: Conceptualization, Investigation.
Xueju Wang: Conceptualization, Dataset.
Lele Cong: Conceptualization, Dataset.
Xianling Cong: Conceptualization, Dataset.
Shisong Tang: Conceptualization, Dataset.
Hechang Chen: Conceptualization, Dataset.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5033167